ตัวแบบการเรียนรู้จำแนกประเภทซัพพลายเออร์แบบมีผู้สอนสำหรับปัญหาการประเมินประสิทธิภาพของซัพพลายเออร์ในระบบ SAP ERP

ณรงค์ศักดิ์ โค้ววิลัยแสง[1]
สาขาวิชาการวิเคราะห์ธุรกิจและวิทยาการข้อมูล คณะสถิติประยุกต์ สถาบันบัณฑิตพัฒนบริหารศาสตร์
และ อัครนันท์ พงศธรวิวัฒน์[2*]
สาขาวิชาการจัดการโลจิสติกส์ คณะสถิติประยุกต์ สถาบันบัณฑิตพัฒนบริหารศาสตร์
148 หมู่ 3 ถนนเสรีไทย แขวงคลองจั่น เขตบางกะปิ กรุงเทพฯ 10240

Received: 11 April 2021; Revised: 13 May 2021; Accepted: 18 May 2021

บทคัดย่อ

งานวิจัยนี้มีวัตถุประสงค์เพื่อพัฒนากรอบแนวคิดการจำแนกประเภทซัพพลายเออร์สำหรับปัญหาการประเมินซัพพลายเออร์ในระบบ SAP ERP ที่มีความไม่แน่นอนจากการวิเคราะห์ค่าน้ำหนักเชิงเส้น (Linear scoring model) โดยนำรายการสั่งซื้อ (Purchase orders) มาสร้างแบบจำลองการแยกประเภทซัพพลายเออร์ และข้อมูลใบขอซื้อ (Purchase requisition) เป็นข้อมูลนำเข้าเพื่อจำแนกประสิทธิภาพซัพพลายเออร์ก่อนตัดสินใจสั่งซื้อสินค้ากับซัพพลายเออร์นั้น ๆ เกณฑ์ในการประเมินซัพพลายเออร์พัฒนาจากการสัมภาษณ์ผู้เชี่ยวชาญ ประกอบด้วยจำนวนสินค้า (Quantity) คุณภาพของสินค้า (Quality) ความน่าเชื่อถือในการจัดส่งสินค้า (Delivery commitment) รวมถึงการพัฒนาขั้นตอนวิธีการดึงข้อมูล (Extract) จากระบบ SAP ERP และแปลงข้อมูล (Transform) ให้สามารถนำมาคำนวณด้วยแบบจำลองจำแนกประเภทประสิทธิภาพของซัพพลายเออร์ทั้งข้อมูลในการเรียนรู้และข้อมูลทดสอบตามระดับการวิเคราะห์ภาพรวมทุกกลุ่มสินค้า (All material groups) และระดับกลุ่มสินค้า (Material group) โดยทำการเปรียบเทียบแบบจำลอง 9 แบบคือ ตัวแบบการจำแนกประเภทแบบเบย์ส (Naïve bay) ตัวแบบการจำแนกประเภท (K-Nearest Neighbors) ตัวแบบจำแนกประเภทซัพพอร์ทเวกเตอร์-แมชชีน (Support Vector Machine) ตัวแบบการถดถอยลอจีสติก (Logistic Regression) ตัวแบบจำแนกประเภทเอดาบูท (Adaboost) ตัวแบบจำแนกประเภทต้นไม้ตัดสินใจ (Decision Tree) ตัวแบบจำแนกประเภทการสุ่มป่าไม้ (Random Forest) และการร่วมกันตัดสินใจสองตัวแบบ คือ Stacking SVM+DT และ Stacking SVM+LR ผลการศึกษาพบว่า แบบจำลองที่ดีที่สุดทั้งในระดับการวิเคราะห์ภาพรวมทุกกลุ่มสินค้าและระดับการวิเคราะห์กลุ่มสินค้า คือ ตัวแบบจำแนกประเภทเอดาบูท Adaboost ที่มีค่าความแม่นยำสูงที่สุดอยู่ที่ร้อยละ 81.6 และ 67.6 ตามลำดับ

คำสำคัญ: การประเมินซัพพลายเออร์ การจำแนกซัพพลายเออร์ การเรียนรู้ของเครื่องจักร กรณีศึกษา SAP ERP

* Corresponding author. E-mail: ckhomp@gmail.com
[1] สาขาวิชาการวิเคราะห์ธุรกิจและวิทยาการข้อมูล คณะสถิติประยุกต์ สถาบันบัณฑิตพัฒนบริหารศาสตร์
[2] สาขาวิชาการจัดการโลจิสติกส์ คณะสถิติประยุกต์ สถาบันบัณฑิตพัฒนบริหารศาสตร์

A Supervised Learning-Based Supplier Classification Model for Supplier Performance Evaluation Problem in SAP ERP

Narongsak Kowwilaisaeng[1]

Business Analytics and Data Science, National Institute of Development Administration (NIDA)

and Akkaranan Phongsatornwiwat[2*]

Logistics Management, National Institute of Development Administration (NIDA),

148 Moo 3, Seri-thai Road, Klong-Chan, Bangkapi, Bangkok Thailand 10240

Received: 11 April 2021; Revised: 13 May 2021; Accepted: 18 May 2021

Abstract

The purpose of this research is to develop a supplier classification framework for supplier evaluation problem for overcoming the uncertain results from embedded linear scoring methods in SAP ERP. Datasets in this study are Purchase Order (PO) and Purchase Requisition (PR). PO data is used for modelling supplier classification models, and PR is used as an input data to evaluate the suppliers' performances before making purchasing contracts. Three evaluation criteria, which are developed by interviewing with experts are quantity, quality, and delivery commitment. In addition, this research is developed the data extraction algorithm from SAP ERP system and data transformation in order for designing suitably datasets for supplier classification models. Nine classification models: Naïve bay, K-Nearest Neighbors, Support Vector Machine, Logistic Regression, Adaboost, Decision Tree, Random Forest, and Stacking SVM+DT and Stacking SVM+LR are applied to compare the performances to find the best classification model. In the analysis schemes, all material groups and each material group as suggested two levels of analysis are proposed in this study. Results show that Adaboost is the best classification model in both train and test datasets. Furthermore, Adaboost is also outperform others for both levels of analysis with accuracy performances for all material group (81.6%) and each material group (67.6), respectively.

Keywords: supplier evaluation, supplier classification, machine learning, case study, SAP ERP

* Corresponding author. E-mail: akkaranan@as.nida.ac.th

[1] Business Analytics and Data Science, National Institute of Development Administration

[2] Logistics Management, National Institute of Development Administration

1. บทนำ

กระบวนการจัดซื้อจัดจ้างเป็นส่วนสำคัญมากในการขับเคลื่อนห่วงโซ่อุปทานและเป็นหนึ่งในกลยุทธ์ในการเพิ่มขีดความสามารถในการแข่งขันทางธุรกิจ [1] ส่งผลให้กิจกรรมการจัดซื้อและการประเมินการเลือกซัพพลายเออร์มีความสัมพันธ์กันอย่างมากกับประสิทธิภาพขององค์กร การเลือกซัพพลายเออร์ที่เหมาะสมจึงเป็นข้อกำหนดเบื้องต้นที่สำคัญสำหรับองค์กร ซึ่งจะนำไปสู่การบรรลุกลยุทธ์การแข่งขันขององค์กรได้ดียิ่งขึ้น [2]

ธุรกิจขนาดกลางไปจนถึงขนาดใหญ่นิยมใช้ระบบสารสนเทศในการบริหารจัดการ เช่น SAP ERP ในการจัดการฟังก์ชันการจัดซื้อจัดจ้างเพื่อให้การจัดการมีประสิทธิภาพมากยิ่งขึ้น เพราะข้อดีของระบบ SAP ERP คือ สามารถจัดการได้หลายส่วนงาน เช่น การวางแผนการจัดซื้อสินค้า การจัดการสินค้าคงคลัง การจัดการผู้ขาย เป็นต้น [3] นอกจากนั้น ระบบ SAP ERP ยังมีแอฟพลิเคชันที่สนับสนุนกระบวนการคัดเลือกและประเมินซัพพลายเออร์ (suppliers) อีกด้วย ซึ่งวิธีการคำนวณนั้น จะใช้หลักการของการคำนวณคะแนนเชิงเส้น (linear score model) [4] อย่างไรก็ดี ข้อจำกัดที่สำคัญของวิธีการดังกล่าวนั้นคือ การกำหนดค่าระดับความสำคัญของปัจจัย (weight importance) ที่ขึ้นอยู่กับประสบการณ์ของผู้ให้คะแนน ทำให้ผลการประเมินผลซัพพลายเออร์นั้นเกิดความไม่แน่นอนขึ้น เพราะระดับคะแนนขึ้นอยู่กับผู้ประเมิน [4, 10, 13] โดยเฉพาะอย่างยิ่งการพยากรณ์หรือการจำลองการเลือกซัพพลายเออร์จากประวัติการทำรายการข้อมูลของซัพพลายเออร์ในอดีต [10] ทำให้เป็นอีกหนึ่งข้อจำกัดของระบบ SAP ERP

ปัญหาการคัดเลือกซัพพลายเออร์นั้นเป็นปัญหาการตัดสินใจตัดสินใจแบบหลายหลักเกณฑ์ (Multiple Criteria Decision Making: MCDM) โดยปัจจัยในการประเมินมีทั้งเชิงคุณภาพและเชิงปริมาณ [1-2, 4,13] ซึ่งวิธีการวิจัยหลายปีที่ผ่านมามีการใช้วิธีทางสถิติและคณิตศาสตร์ เช่น การวิเคราะห์เชิงเส้นโอบล้อม Data Envelopment Analysis (DEA) [1]ม การวิเคราะห์เชิงลำดับชั้น (Analytics Hierarchy Process: AHP) [2], 2-Tuple decision model [4] และ Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS) [7] จากการทบทวนวรรณกรรมที่เกี่ยวข้องพบว่า ปัญหาการคัดเลือกซัพพลายเออร์นั้น ประกอบไปด้วยสองกระบวนการ คือ การกำหนดปัจจัยในการคัดเลือก (Evaluation Criteria) และการพัฒนาตัวแบบการตัดสินใจ (Decision-Making Model) ทั้งนี้ตัวแบบการตัดสินใจจะขึ้นอยู่กับปัจจัยในการคัดเลือก [4,13] อย่างไรก็ดี วิธีการดังกล่าวจะเป็นตัวแบบแบบคงที่ (Deterministic Models) ถ้ามีการปรับเปลี่ยนคะแนนหรือค่าน้ำหนักปัจจัยต้องมีการปรับตัวแบบอีกครั้งให้เหมาะสมกับสถานการณ์

การเรียนรู้ของเครื่องจักร หรือ Machine Learning (ML) จึงเป็นอีกทางเลือกหนึ่งในปัญหาการประเมินซัพพลายเออร์ การเรียนรู้ของเครื่องจักรสามารถแก้ปัญหาที่มีความซับซ้อนโดยใช้ข้อมูลในอดีตในการเรียนรู้ [3] ทำให้สามารถตัดปัญหาในเรื่องความไม่แน่นอนที่เกิดจากความไม่รู้ (Ignorance) จากประสบการณ์ของผู้ประเมินได้ [4-6] เทคนิคที่นำมาประยุกต์ใช้ในปัญหาการทำนายกลุ่มซัพพลายเออร์ (Supplier Classification) นั้นประกอบไปด้วยตัวแบบจำแนกประเภทซัพพอร์ทเวกเตอร์แมชชีน (Support Vector Machine: SVM) [10], ตัวแบบจำแนกแบบแผนภูมิต้นไม้ (Decision Tree: DT) [10] เป็นอัลกอริทึมที่สำคัญของการเรียนรู้ด้วยเครื่องจักรที่สามารถนำไปประยุกต์ใช้ในการแก้ปัญหาการจำแนกประเภทซัพพลายเออร์ได้ [8] ได้ทำการประยุกต์ใช้เทคนิค AHP กับเทคนิคการเรียนรู้ของเครื่องจักรเพื่อทำการพัฒนาระบบในการแนะนำการเลือกซื้อเครื่องดื่มประเภทไวน์ (Wine Recommendation System) [9] ได้ทำการศึกษาระดับความสำคัญของปัจจัย (Feature Importance) ที่จะนำมาประเมินประสิทธิภาพของซัพพลายเออร์ โดยการนำเทคนิคการเรียนรู้ของเครื่องจักรเข้ามาช่วยในการหาปัจจัยทีมีค่าความสำคัญ พบว่า การส่งมอบที่ตรงเวลาเป็นตัวบ่งชี้ความน่าเชื่อถือและประสิทธิภาพของซัพพลายเออร์ [14] ได้ทำการพัฒนาตัวแบบการประเมินซัพพลายเออร์ด้วยเทคนิคการเรียนรู้แบบมีผู้สอน โดยเทคนิคที่พัฒนาขึ้นนั้นพัฒนาต่อยอดจากตัวแบบซัพพอร์ทเวกเตอร์แมชชีน (Support Vector Machine) และ Ian M. Cavalcante [8] ได้พัฒนาระบบ

อัตโนมัติด้วยตัวแบบการประเมินซัพพลายเออร์แบบมีผู้สอน เพื่อคัดเลือกซัพพลายเออร์ที่มีประสิทธิภาพที่ดีที่สุดในอุตสาหกรรมการผลิตแบบดิจิตอล และ Manu Kohli [10] ได้ทำการพัฒนาระบบการคัดเลือกซัพพลายเออร์ด้วยเทคนิคการเรียนรู้ของเครื่องจักร ข้อมูลที่นำเข้าสอนระบบนั้นเป็นใบสั่งซื้อสินค้า (Purchase Orders) ในระบบ SAP ERP ซึ่งทำการพัฒนาคำสั่งเพื่อดึงข้อมูลมาทำการวิเคราะห์ ผลการศึกษาพบว่า DT และ SVM เป็นตัวแบบที่ดีที่สุดในการจำแนกประสิทธิภาพของซัพพลายเออร์ อย่างไรก็ตามยังไม่ได้มีการทดสอบกับชุดข้อมูลใหม่ เช่น ใบคำสั่งต้องการสินค้า (Purchase Requisitions)

เพื่อทำการแก้ไขข้อจำกัดดังกล่าว งานวิจัยนี้มีวัตถุประสงค์เพื่อพัฒนากรอบแนวคิดของตัวแบบการจำแนกประเภทซัพพลายเออร์สำหรับปัญหาการประเมินซัพพลายเออร์ในระบบ SAP ERP ซึ่งใช้เทคนิคการเรียนรู้ของเครื่องจักรในการจัดจำแนกประเภท ด้วยข้อมูลการทำรายการสั่งซื้อ (Purchase Orders) ในอดีตจากระบบ SAP ERP และนำมาทดสอบประสิทธิภาพของระบบด้วยข้อมูลใบคำสั่งต้องการสินค้า (Purchase Requisitions) ประโยชน์ของงานวิจัยนี้คือ สามารถช่วยเพื่อความสามารถของผู้ประกอบการธุรกิจขนาดกลางไปจนถึงขนาดใหญ่ในการประเมินซัพพลายเออร์ก่อนการตัดสินใจสั่งซื้อสินค้ากับซัพพลายเออร์นั้น ๆ ได้อย่างมีประสิทธิภาพมากขึ้น และยกระดับศักยภาพในการดำเนินการธุรกิจเพื่อการแข่งขันได้เพิ่มมากขึ้น

2. งานวิจัยและทฤษฎีที่เกี่ยวข้อง

การศึกษาการประเมินซัพพลายเออร์ที่ผ่านมาสามารถแบ่งออกได้เป็น 6 ประเภทสำคัญ [7] คือ สถิติและความน่าจะเป็น(Statistics/Probabilistic), การตัดสินใจแบบหลายคุณลักษณะ (Multi-Attribute Decision Making), วิธีการวิเคราะห์ด้วยการพิจารณาต้นทุน (Method-based on costs), โปรแกรมคณิตศาสตร์เชิงเส้น (Mathematical Programming), ปัญญาประดิษฐ์(Artificial Intelligence) และการผสมผสานวิธี (Combined Approaches) โดยเทคนิคที่ได้รับความนิยมสำหรับปัญหาการคัดเลือกซัพพลายเออร์ (Supplier Selection) นั้น ประกอบไปด้วยการตัดสินใจแบบหลายคุณลักษณะ และ ปัญญาประดิษฐ์ (Artificial Intelligence)

การประเมินผลและการคัดเลือกซัพพลายเออร์ด้วยปัญญาประดิษฐ์ สามารถช่วยแก้ปัญหากระบวนการที่ซับซ้อนในการตัดสินใจแบบหลายหลักเกณฑ์ [10] เพราะคุณลักษณะของซัพพลายเออร์ เช่น ความสามารถของซัพพลายเออร์แต่ละรายนั้น บางครั้งมีความแตกต่างกันเพียงเล็กน้อย เป็นสิ่งที่ยากต่อการตัดสินใจ [9] จากการทบทวนวรรณกรรมที่เกี่ยวข้อง พบว่า การประยุกต์ใช้ปัญญาประดิษฐ์มีประสิทธิภาพดีกว่าวิธีทั่วไปในการพิจารณาหาซัพพลายเออร์ที่ดีที่สุด [10-16] ตัวอย่างเช่น ตัวแบบต้นไม้ตัดสินใจ (Decision Tree: DT) เป็นวิธีการแบบกราฟิกที่จะแสดงตัวเลือกที่แตกต่างกันทำโดยบุคคลหรือเครื่องจักรในรูปแบบของต้นไม้ โหนด (node) ของกราฟแสดงให้เห็นถึงการเกิดขึ้นของเหตุการณ์และขอบแสดงกระบวนทัศน์การตัดสินใจ วิธีนี้คือใช้ในการเรียนรู้ของเครื่องและการทำเหมืองข้อมูล [9] เรียนรู้ที่จะจำแนกซัพพลายเออร์ออกเป็น 2 ประเภท: ดีและไม่ดี นอกจากนี้แบบจำลองการแยกประเภท DT ยังใช้เป็นเกณฑ์สำคัญเพื่อลดความซับซ้อนของกระบวนการ AHP [5] นอกจากนั้น Support Vector Machines (SVMs) ก็เป็นอีกหนึ่งอัลกอริทึมการเรียนรู้ของเครื่องจักรที่ใช้สำหรับการจำแนกประเภท สามารถใช้เพื่อจำลองประสิทธิภาพของซัพพลายเออร์ในระบบบริหารความสัมพันธ์คู่ค้าได้เมื่อเทียบกับการจำแนกประเภทเชิงเส้นแบบดั้งเดิมมีความสามารถในการสร้างแบบจำลองได้มีประสิทธิภาพดีกว่า เนื่องจากสามารถใช้การผสมผสานข้อมูลนำเข้าแบบไม่เชิงเส้นได้ [14] เห็นได้ว่าแนวคิดเกี่ยวกับประสิทธิภาพของซัพพลายเออร์ สามารถผสมผสานกับเทคนิคการเรียนรู้และการจำลองด้วยเครื่องจักร เพื่อนำมาวิเคราะห์เพื่อแก้ปัญหาการคัดเลือกซัพพลายเออร์ได้ [8] ในส่วนต่อไปจะนำเสนอแนวคิดของเทคนิคการเรียนรู้ของเครื่องจักแต่ละรูปแบที่นำมาปรับใช้ในงานวิจัยครั้งนี้

2.1 การจำแนกประเภทแบบเบย์ (Naïve Bays)

ตัวแบบที่คำนวณค่าความน่าจะเป็นที่ข้อมูลจะมีประเภทใดประเภทหนึ่ง ข้อมูลที่กำหนดมีประเภทข้อมูลที่ให้ค่าความน่าจะเป็นสูงสุด [3]

$$P(c|x) = \frac{P(x|c)P(c)}{P(c)} \quad (1)$$

$$\mathrm{P(c|x)} = \mathrm{P(x_1|c)} \times \mathrm{P(x_2|c)} \times \ldots \times \mathrm{(x_n|c)} \times \mathrm{P(c)} \quad (2)$$

เมื่อ c คือ ค่าเป้าหมาย

x คือ ตัวแปรอิสระ

และ P คือ ค่าความน่าจะเป็น

เนื่องจากค่าความน่าจะเป็นของทุกประเภทมีตัวหารที่เป็นค่าตัวเดียวกัน การหาค่า c ที่ให้ความน่าจะเป็นตามสมการสูงสุด สามารถทำได้โดยการคำนวณเฉพาะค่าผลคูณที่เป็นตัวตั้งของสมการแล้วนำค่าผลคูณของแต่ละประเภทมาเปรียบเทียบ เพื่อหาค่าสูงสุดสำหรับ $P(c)$ เป็น Prior Probability [3]

2.2 ต้นไม้ตัดสินใจ (Decision Tree)

ต้นไม้ตัดสินใจ (Decision Tree) เป้าหมายของแบบจำลองคือต้องการจำแนกประเภทได้ถูกต้องมากที่สุด ความลึกของ node root ไปจนถึง node leaf ที่ไกลที่สุดให้มีค่าน้อยที่สุด โดยการสร้างเงื่อนไขการตัดสินใจสำหรับแต่ละประเภทตัวแปร [3]

1. Nominal เงื่อนไขตัวแปรเท่ากับค่าที่เป็นไปได้ของตัวแปรนั้น
2. Binary เงื่อนไขเป็นไปได้ 2 ทาง กลุ่มใด กลุ่มหนึ่ง
3. Ordinal เงื่อนไขตัวแปรทำได้แบบเดียวกับตัวแปรประเภท Nominal แต่ข้อมูลทั้งหมดต้องเรียงลำดับหรือการจัดกลุ่ม อย่างชัดเชนเช่น {เล็ก, กลาง, ใหญ่}, {ใหญ่, กลาง}, {กลาง, เล็ก}
4. Numerical เงื่อนไขตัวแปรแบ่งข้อมูลเป็นช่วง (Interval) หรือแบ่งความกว้างเท่า ๆ กัน (Equal width)

หลักการในการแบ่งข้อมูลแต่ละ node สำหรับข้อมูลที่มี k feature และ n observation โดยการวัดความไม่บริสุทธิ์ของข้อมูล (Impurity) เพื่อต้องการ minimize impurity ตัวชี้วัดบอกระดับความไม่บริสุทธิ์ Gini Index ยิ่งน้อยยิ่งดี หรือ Entropy วัดความไม่แน่นอน (randomness) ของข้อมูล

$$Gini(t) = 1 - \sum_{j}^{n} P^2(j) \quad (3)$$

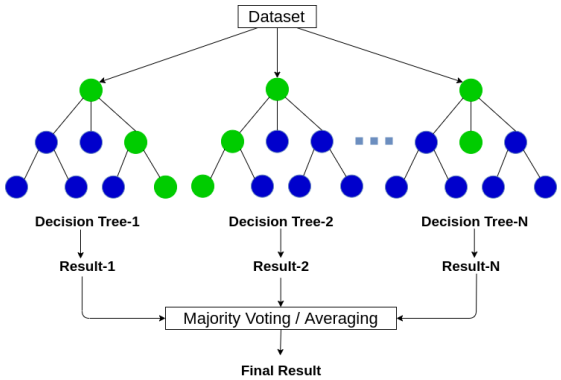
$$Entropy = -\sum_{j}^{n} P(j)\log\left(P(j)\right) \quad (4)$$

เมื่อ $P(j)$ คือ สัดส่วนของข้อมูลประเภท j

และ n คือ จำนวนประเภทของข้อมูล

2.3 การสุ่มป่าไม้ (Random Forest)

การสุ่มป่าไม้ (Random Forest) เป็นหนึ่งในกลุ่มของแบบจำลองแยกประเภทที่เรียกว่า Ensemble learning ที่มีหลักการคือการเรียนรู้ของแบบจำลองที่เหมือนกันหลายๆ ครั้ง (หลาย Instance) บนข้อมูลชุดเดียวกัน [8] โดยแต่ละครั้งของการเรียนรู้จะเลือกส่วนของข้อมูลที่เรียนรู้ไม่เหมือนกัน แล้วเอาการตัดสินใจของแบบจำลองแยกประเภทเหล่านั้นมาโหวต (Majority vote) กันว่า Class ไหนถูกเลือกมากที่สุดหลักการตัดสินใจดังรูปที่ 1



รูปที่ 1 หลักการตัดสินใจของ Random Forest (https://www.analyticsvidhya.com/blog/2020/05/decision-tree-vs-random-forest-algorithm/

2.4 ซัพพอร์ตเวกเตอร์ (Support Vector Machine)

ซัพพอร์ตเวกเตอร์ (Support Vector Machine) เป็นตัวแบบจำแนกประเภทข้อมูลโดยใช้ Hyperplane ในการแบ่งข้อมูลใน Multidimensional Space โดยสมมติฐานว่าข้อมูลมีคุณสมบัติเป็น Linear Separable กล่าวคือ ข้อมูลทั้งหมดที่มีตำแหน่งอยู่เหนือ Hyperplane จะเป็นประเภทหนึ่งและข้อมูลทั้งหมดที่อยู่ใต้ Hyperplane จะเป็นอีกประเภทหนึ่ง

$$Maximize: \sum_{i=1}^{n} \alpha_i - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j (u_i \cdot u_j) \quad (5)$$

เมื่อ n คือ จำนวนข้อมูล n จำนวน (u_1,y_1), (u_2,y_2), ... , (u_n,y_n)

u คือ เวกเตอร์แทนข้อมูลเรียนรู้ใด ๆ

y = 1 เมื่อข้อมูล u มีประเภทเป็นบวก

y = 0 เมื่อข้อมูล u มีประเภทเป็นลบ

α_i คือ Lagrange Multipliers ของข้อมูลแต่ละตัวในชุดเรียนรู้

และ $i = 1, 2, \ldots, n$

โดยปกติข้อมูลอาจมีการกระจายตัวที่การแบ่งข้อมูลทั้งสองประเภทออกจากกันด้วย Hyperplane ไม่สามารถทำได้ การใช้ตัวแบบ SVM ในการจำแนกประเภทจึงทำไม่ได้โดยตรง [3] วิธีการแก้ไขก็คือการแปลงข้อมูลที่อยู่ใน input space ไปสู่ transformed Space ที่เรียกว่า Feature space จึงสามารถใช้ตัวแบบ SVM ในการแบ่งข้อมูลด้วย Hyperplane เช่น Polynomial Kernel, Radial Basis Function(RBF), Sigmoid Kernel การแยกประเภทข้อมูลที่มีมากกว่า 2 ประเภท สามารถทำได้โดยสร้างตัวแบบ SVM สำหรับจำแนกประเภทข้อมูล 2 ประเภทหลายๆ ตัวมาตัดสินใจร่วมกัน มักนิยมทำใน 2 ลักษณะ [3]

1. One-Versus-Rest เป็นการสร้างตัวแบบ SVM แบบ 2 ประเภทประเภทหนึ่ง (Positive Class) กับข้อมูลอื่นที่เหลือ (Negative Class)
2. One-Versus-One เป็นการสร้างตัวแบบ SVM แบบ 2 ประเภท แต่ละตัวจะจำแนกข้อมูลระหว่างแต่ละคู่ของประเภทข้อมูล ถ้ามีข้อมูล n ประเภทจะต้องสร้างตัวแบบ SVM จำนวน n(n-1)/2 ตัว

2.5 เพื่อนบ้านใกล้ที่สุด (K-Nearest Neighbors)

การสร้างแบบจำลองจำแนกประเภทของข้อมูลโดยการเปรียบเทียบข้อมูลตัวอย่างที่เก็บอยู่ในฐานข้อมูล (Instance-based Learning) จะค้นหาตัวอย่างข้อมูลในฐานที่มีลักษณะใกล้เคียงข้อมูลใหม่มากที่สุดจำนวน K ตัว มาร่วมกันตัดสินว่าข้อมูลใหม่นี้ควรเป็นประเภทใด [3] โดยใช้หลักการตัดสิน Majority vote การวัดความเหมือนระหว่างข้อมูลใหม่และข้อมูลในฐานข้อมูลสามารถใช้ตัววัดความเหมือน (Similarity) เช่น Euclidean distance เป็นต้น

$$d = \sqrt{\sum_{i=1}^{n}(p_i - q_i)^2} \quad (6)$$

เมื่อ p คือ ค่าของข้อมูลที่ต้องการจำแนก

q คือ ค่าของชุดข้อมูลเพื่อนบ้านที่นำมาพิจารณา

2.6 การวิเคราะห์ถดถอยลอจิสติก (Logistic Regression)

การวิเคราะห์การถดถอยลอจิสติกแบ่งออกเป็น 2 ประเภท คือ (1) Binary logistic regression เป็นการวิเคราะห์ความสัมพันธ์ระหว่างตัวแปรตาม 1 ตัวกับตัวแปรต้นหลายตัว เพื่อแสดงความน่าจะเป็นของสิ่งที่สนใจ เช่น ดี/ไม่ดี, ใช่/ไม่ใช่ และ (2) Multinomial logistic regression เป็นการวิเคราะห์ตัวแปรที่มีมากกว่า 2 กลุ่มขึ้นไป สถิติที่ใช้เปรียบเทียบค่า odd Ratio รูปแบบสถิติพัฒนามาจากฟังก์ชันลอจิสติกฟังก์ชัน (Logistic function) ซึ่งเป็นฟังก์ชันที่แสดงโอกาสที่จะเกิดและไม่เกิดเหตุการณ์ในรูปแบบสมการเอกโปเนนเชียล (Exponential) [3]

$$Odd\ Ratio = \frac{p}{1-p} \quad (7)$$

$$Logit(p) = \log\left(\frac{p}{1-p}\right) \quad (8)$$

เมื่อ p คือ ค่าความน่าจะเป็นที่จะเกิดเหตุการณ์

2.7 การทำงานร่วมกัน หรือ เอ็นเซมเบิล (Ensemble Method)

การสร้างตัวแบบจำแนกประเภทข้อมูลหลาย ๆ ตัวมาร่วมกันตัดสินประเภทข้อมูล โดยผลลัพธ์ของการตัดสินร่วมกันของตัวแบบจำแนกประเภทข้อมูลเหล่านี้ จะให้ความแม่นยำกว่าการใช้ตัวแบบจำแนกประเภทข้อมูลเพียงตัวเดียว [3] โดยใช้หลักการ Majority vote ในการตัดสินใจขั้นสุดท้าย การสร้างตัวแบบหลายตัวเพื่อร่วมการตัดสินประเภทข้อมูลสามารถทำได้ใน 2 ลักษณะ

1. Bagging เป็นวิธีการที่ตัวแบบแต่ละตัวใน Ensemble method จะถูกสร้างจากชุดข้อมูลที่สุ่มจากชุดข้อมูลเรียนรู้ที่กำหนดให้แบบคืน (Sampling With Replacement) มีเป้าหมายที่จะลดความผิดพลาดจากข้อมูลที่มีความจำเพาะมากและมีซับซ้อนสูง ซึ่งหมายถึงอาจมีสภาวะ Overfitting

2. Boosting เป็นวิธีที่ตัวแบบแต่ละตัวของ Ensemble method ถูกสร้างโดยใช้ชุดข้อมูลที่สุ่มมาจากชุดข้อมูลเรียนรู้ที่กำหนดให้แบบคืน เช่นเดียวกับ Bagging ข้อแตกต่างคือ การสุ่มข้อมูลแล้วสร้างตัวแบบไปทีละตัวในแต่ละรอบการทำงาน โดยภายหลังแต่ละรอบจะมีการปรับน้ำหนักให้โอกาสในการสุ่มข้อมูลแต่ละตัวในชุดข้อมูลการเรียนรู้มากขึ้นในรอบถัดไป เป้าหมายที่จะลดความผิดพลาดจากข้อมูลที่มีความจำเพาะมากเกินไป ซึ่งหมายถึงอาจมีสภาวะ Overfitting ยังช่วยความผิดพลาดจากข้อมูลที่มีความจำเพาะน้อยด้วย ซึ่งหมายถึง สภาวะ Underfitting

2.8 เอดาบูท (Adaboost)

Adaboost หรือ adaptive boosting เป็นการเรียนรู้ต่อกันเป็นลูกโซ่ (Sequential ensemble method) ที่รวม weak learners หลาย ๆตัวแล้วสร้างเป็น strong learning ทำให้ประสิทธิภาพของแบบจำลองจำแนกประเภทมีประสิทธิภาพเพิ่มขึ้น [3] โดยการเรียนรู้ในแต่ละรอบจะมีการกำหนดค่าน้ำหนักของข้อมูลแต่ละรายการ (Classifier Instance) จะเรียนรู้จากค่าน้ำหนักเหล่านั้น โดยถ้าจำแนกประเภทผิด ค่าน้ำหนักของรายการนั้นจะมีค่าน้ำหนักมากขึ้น การจำแนกประเภทของ Adaboost คือการถ่วงคะแนนเข้ากับคำตอบของแต่ละ instance แล้วเลือกคำตอบจากค่าเป้าหมาย (Class) ที่ได้รวมแล้วค่าน้ำหนักมากที่สุด [3] สังเกตว่า AdaBoost ให้ค่ากับ Instance ที่ "เก่ง" มากกว่าที่ "ไม่เก่ง" ในขณะที่ Ensemble learning แบบ Random forest ให้ค่าน้ำหนักทุก Instance เท่ากัน AdaBoost ไม่ได้มีเป้าหมายที่ทำให้ Classifier instance ตัวถัดๆ ไปมีความแม่นยำขึ้น แต่ทำงานโดยการเก็บคะแนนความแม่นยำของแต่ละ Instance เอาไว้เพื่อเอามาถ่วงน้ำหนักการตัดสินใจเมื่อต้องการพยากรณ์ในภายหลัง

$$H(x) = \text{sign}\left(\sum_{t=1}^{T} \alpha_t h_t(x)\right) \quad (9)$$

เมื่อ H คือ Strong learner

t คือ จำนวนรอบการเรียนรู้ t = 1, 2, 3, ..., T

α_t คือ voting power

และ $ht(x)$ คือ Weak learner

2.9 การประเมินตัวแบบการจำแนกประเภทข้อมูล

การประเมินตัวแบบจำแนกประเภทข้อมูล เป็นขั้นตอนสำคัญ เพื่อให้ทราบถึงประสิทธิภาพของตัวแบบที่ถูกสร้างขึ้นมารวมทั้งสามารถเปรียบเทียบตัวแบบเพื่อเลือกตัวแบบที่ดีที่สุดในการนำไปใช้ [3] การวัดประสิทธิภาพการจำแนกประเภทสามารถใช้เครื่องมือหรือตัวชี้วัดดังนี้

Confusion Matrix จะเป็นตารางแสดงจำนวนของข้อมูลในแต่ละประเภทที่ถูกจำแนกด้วยตัวแบบ [3] ดังตารางที่ 1

ตารางที่ 1 Confusion Matrix

		Predicted Class	
	Class	Positive	Negative
Actual Class	Positive	TP	FN
	Negative	FP	TN

True Positive (TP) เป็นจำนวนข้อมูลที่มีประเภทเป็น Positive และถูกตัวแบบจำแนกประเภทเป็น Positive

False Positive (FP) เป็นจำนวนข้อมูลที่มีประเภทเป็น Negative และถูกตัวแบบจำแนกประเภทเป็น Positive

True Negative (TN) เป็นจำนวนข้อมูลที่มีประเภทเป็น Negative และถูกตัวแบบจำแนกประเภทเป็น Negative

False Negative (FN) เป็นจำนวนข้อมูลที่มีประเภทเป็น Positive และถูกตัวแบบจำแนกประเภทเป็น Negative

ตัวชี้วัดจาก Confusion Matrix ได้แก่

1. Accuracy คือ ความแม่นยำในการจำแนกประเภทข้อมูล

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (10)$$

2. Precision คือ ความถูกต้องของการจำแนกประเภท ข้อมูล

$$\text{Precision} = \frac{TP}{TP + FP} \quad (11)$$

3. Recall คือ ความสามารถของตัวแบบในการจำแนกประเภท

$$\text{Precision} = \frac{TP}{TP + FN} \quad (12)$$

4. F-Measure

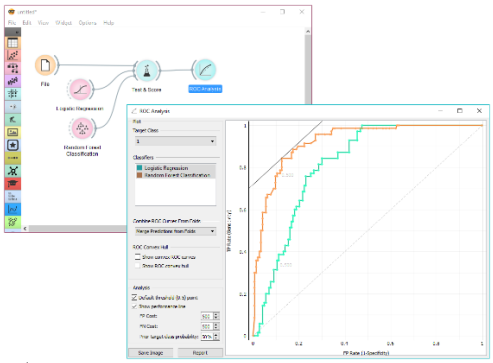
$$\text{F} - \text{Measure} = \frac{2 * Recall * Precision}{Recall + Precision} \quad (13)$$

2.10 Cross Validation (CV)

Cross Validation เป็นการแบ่งข้อมูลออกเป็น K Disjoint Subsets (แต่ละ Subsets มีข้อมูลไม่ซ้ำกัน) แล้วใช้ K-1 Subsets ในการสร้างตัวแบบและ 1 Subsets ที่เหลือจะเป็นข้อมูลทดสอบ ทำซ้ำเช่นนี้ K ครั้ง ในแต่ละครั้งจะใช้ K-1 Subsets เป็นข้อมูลเรียนรู้และ 1 Subsets เป็นข้อมูลทดสอบที่ไม่เหมือนกัน การประเมินประสิทธิภาพของตัวแบบด้วยวิธีนี้จะใช้ค่าเฉลี่ยของการประเมินผลทั้ง K ครั้ง วิธีนี้เรียกว่า K-fold Cross validation [3]

2.11 Orange Data mining

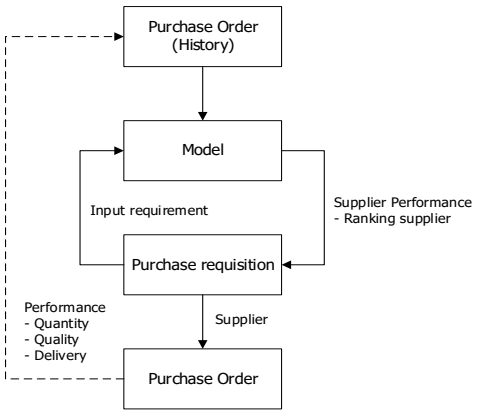
Orange Data mining เป็นซอฟแวร์โอเพนซอร์สสำหรับการเรียนรู้ของเครื่องจักรและแสดงผลข้อมูล สำหรับการสร้างกระบวนการวิเคราะห์ข้อมูล (Data Analytics workflow) รูปแบบการใช้เป็นลักษณะการเขียนโปรแกรมด้วยภาพ (Visual Programming) มุ่งเน้นการวิเคราะห์ข้อมูลแทนการเขียนโค้ด ทำให้การสร้างต้นแบบการวิเคราะห์ข้อมูลได้อย่างรวดเร็ว [11]



รูปที่ 2 ตัวอย่างหน้าจอโปรแกรม Orange Data Mining (https://orangedatamining.com/screenshots/)

3. วิธีการดำเนินงานวิจัย

การศึกษาในครั้งนี้ต้องการนำข้อมูลใบสั่งซื้อ (Purchase Order) ในอดีตมาทำการสร้างแบบจำลองการจำแนกประเภทประสิทธิภาพของซัพพลายเออร์ โดยทำการสร้างแบบจำลองดังนี้ 1. Naïve bay: NB, 2. -Nearest Neighbors: k-NN, 3. Support Vector Machine: SVM 4. Logistic Regression: LR, 5. Adaboost 6. Decision Tree: DT, 7. Random Forest: RF, 8. Stacking SVM+DT, 9. Stacking SVM+LR เพื่อคัดเลือกแบบจำลองที่เหมาะสม เมื่อได้แบบจำลองที่เหมาะสมแล้วจะนำมาใช้ในการตัดสินใจในการเลือกซัพพลายเออร์ที่ดีที่สุดดังแสดงในรูปที่ 3



รูปที่ 3 กรอบแนวคิดการจำแนกประเภทซัพพลายเออร์

3.1 ข้อมูลที่ใช้ในการศึกษา

ข้อมูลสำหรับการศึกษาในครั้งนี้ นำมาจากข้อมูล Transactions module MM และ Module QM ในระบบ SAP ERP ของบริษัทอุตสาหกรรมอาหารและเครื่องดื่มแห่งหนึ่งในประเทศไทย เป็นข้อมูลที่อยู่ในระบบตั้งแต่กันยายน 2562 ถึง ตุลาคม 2563 ซึ่งข้อมูลอาจมีความไม่สมบรูณ์สำหรับการประเมินซัพพลายเออร์ เนื่องจากกระบวนการทำงานหรือขั้นตอนการทำงานไม่ได้กำหนดกฎเกณฑ์สำหรับการประเมินซัพพลายเออร์เอาไว้ตั้งแต่ต้น ตัวอย่างข้อมูลสำหรับเรียนรู้ดังรูปที่ 4 จึงต้องพัฒนากฎเกณฑ์การประเมินขึ้น

Vendor	MatGrp	Material	Payment Term	Open Qty	Net Value	Avg.Qty	Avg.Delivey	Avg.Quality	Class
2100000002	IN1F000	IN1F000025	30	400	260000	1.0000	0.8571	1.0000	2
2300000247	IN1F000	IN1F000002	30	200	104000	1.0000	0.8400	1.0000	2
2100000199	IN1F000	IN1F000033	30	25	31250	1.0000	0.4545	1.0000	2
2100000050	IN1F000	IN1F000024	30	40	27600	1.0000	0.6364	1.0000	1
2100000199	IN1F000	IN1F000034	30	25	17500	1.0000	0.4545	1.0000	3
2100000200	IN1F000	IN1F000022	30	100	82000	1.0000	0.7647	1.0000	2
2100000200	IN1F000	IN1F000027	30	50	36000	1.0000	0.7647	1.0000	2
2100000200	IN1F000	IN1F000031	30	200	178000	1.0000	0.7647	1.0000	2
2100000018	IN1F000	IN1F000018	30	4	16000	1.0000	1.0000	1.0000	2
2100000008	IN1F000	IN1F000014	30	20	5000	1.0000	1.0000	1.0000	2
2100000050	IN1F000	IN1F000015	30	100	95000	1.0000	0.6364	1.0000	2
2100000050	IN1F000	IN1F000004	30	20	24000	1.0000	0.6364	1.0000	3

รูปที่ 4 ตัวอย่างข้อมูลสำหรับเรียนรู้

3.2 การเตรียมข้อมูล

จากการสัมภาษณ์เรื่องการประเมินซัพพลายเออร์จากฝ่ายจัดซื้อและผู้เชี่ยวชาญระบบ SAP ERP จำนวน 5 ท่าน โดยพิจารณาคุณสมบัติดังนี้ เป็นผู้นำทางความคิด มีความรอบรู้และประสบการณ์ทำงานด้านการจัดซื้อมากกว่า 10 ปี และทำงานเกี่ยวข้องกับการบริหารงานหรือกระบวนการจัดซื้อในระบบ SAP ERP จากการเก็บรวบรวมด้วยแบบสอบถาม พบว่า เกณฑ์ที่ใช้ในการประเมินประสิทธิภาพของซัพพลายเออร์ที่ผู้เชี่ยวชาญทุกท่านเห็นสอดคล้องกับการทบทวนวรรณกรรมที่ผ่านมา [1-2, 4, 15] โดยให้ความสำคัญกับเกณฑ์ที่ใช้ในการประเมินซัพพลายเออร์ ดังตารางที่ 2

ตารางที่ 2 เกณฑ์ที่ใช้ในการประเมินซัพพลายเออร์

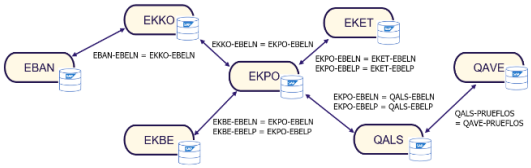
Expert Position/Criteria	Quantit	Quality	Delivery	Price
Purchasing Assistant Manager	*	*	*	*
Senior Purchasing Officer	*	*	*	*
Senior Purchasing Officer	*	*	*	*
SAP Consulting Manager	*	*	*	*
SAP MM Consultant	*	*	*	*

ฐานข้อมูล (Databases) ที่เกี่ยวข้องในการดึงข้อมูลในระบบ SAP ERP ประกอบไปด้วยข้อมูลใบสั่งซื้อและประวัติการเดินรายการจาก Module Material Management (MM) และข้อมูลการให้คะแนนคุณภาพสินค้าจากการควบคุมคุณภาพ Module Quality Management (QM) ดังตารางที่ 3

ตารางที่ 3 ฐานข้อมูลที่เกี่ยวข้องในการดึงข้อมูลในระบบ SAP ERP

Table	Description
EKKO	Purchasing Document Header
EKPO	Purchasing Document Item
EKET	Scheduling Agreement Schedule Lines
EKBE	History per Purchasing Document
EBAN	Purchase Requisition
QALS	Inspection lot record
QAVE	Inspection processing: Usage decision

โดยความสัมพันธ์ของฐานข้อมูลในระบบ SAP ERP โดยเชื่อมโยงกันด้วย Primary key และ Foreign key แสดงดังรูปที่ 5



รูปที่ 5 ความสัมพันธ์ของฐานข้อมูลในระบบ SAP ERP

หลังจากทำการรวบรวมข้อมูลจากฐานข้อมูลที่เกี่ยวข้องแล้ว ทางผู้วิจัยได้ทำการพัฒนาประเภท (Class) ของซัพพลายเออร์ขึ้น โดยปัจจัยที่นำมาพิจารณานั้น ประกอบไปด้วยสามปัจจัยสำคัญ ประกอบด้วย ความสามารถในการส่งสินค้าตามจำนวนที่ต้องการ (Diff. Quantity) ความสามารถในการส่งมอบสินค้าตามเวลาที่กำหนด (Diff. Delivery) และความสามารถในการส่งสินค้าได้ตามรูปแบบที่กำหนด (Quality Score) และผู้วิจัยได้ทำการพัฒนาประเภทของซัพพลายเออร์ขึ้นมาจำนวนทั้งหมด 5 ประเภท ประกอบไปด้วย Class 1 (Excellence) Class 2 (Very good) Class 3 (Good) Class 4 (Satisfactory) และ Class 5 (Unsatisfactory) จากประสบการณ์ของผู้เชี่ยวชาญและประสบการณ์ของฝ่ายจัดซื้อ ดังแสดงในตารางที่ 4

ตารางที่ 4 ฐานข้อมูลที่เกี่ยวข้องในการดึงข้อมูลในระบบ SAP ERP

Evaluation Criteria				
Class	Class Name	Diff Quantity(%)	Diff Delivery(Days)	Quality Score
1	Excellence	< 0	0	100
2	Very good	0	0	100
3	Good	>0 &<10	0	100
4	Satisfactory	>30	>0 &<10	< 100
5	Unsatisfactory	>30	>10	< 100

ตารางที่ 5 ข้อมูลสำหรับการเรียนรู้ของแบบจำลองจำแนกประเภท

Feature	Feature Description
Vendor	รหัสผู้ขาย
Material Group	รหัสกลุ่มสินค้า
Material Number	รหัสสินค้า
Payment Term	เงื่อนไขการชำระเงิน
Open Qty	จำนวนสินค้าจากใบสั่งซื้อสินค้า
Net Value	จำนวนเงิน
Avg.Quantity	ค่าเฉลี่ย การส่งสินค้าครบจำนวน
Avg.Delivey	ค่าเฉลี่ย การส่งสินค้าตรงเวลา
Avg.Quality	ค่าคะแนนเฉลี่ยการตรวจสอบคุณภาพสินค้า
Class	ตัวแปรเป้าหมาย

ขั้นตอนวิธีในการดึงข้อมูลและเตรียมข้อมูล

1 **Data:** Table EKKO (PO Header), EKPO (PO Item), EKET, EKBE(History), QALS, QAVE (Inspection)

Result: Data Model to perform Supplier Evaluation

// Step to extract detail of Purchase Order

2 **While** EKKO-LOEKZ NE 'X' (PO is not deleted)

3 **Do** **If** EKPO-LOEKZ NE 'X' (Item is not deleted)

EKPO-MATNR NE '' and (Material number is not null)

EKPO-KNTTP NE 'A' and EKPO-KNTTP NE 'K' (Account Assignment Category does not Asset and Expense)

4 **Then** Select all entries from Purchase Order line item (EKPO) and PO Header (EKKO)

5 Select all entries form table EKBE

6 Select all entries form table EKET

7 Select all entries form table QALS

8 Select all entries from table QAVE

// Perform criteria Quantity, Quality, Delivery On-Time

9 Total goods received QTY $\sum$ EKBE-MENGE

10 Calculate Over Tolerance EKPO-MENG + (EKPO-MENG * EKPO-UEBTO)/100

11 Calculate Under Tolerance EKPO-MENG + (EKPO-MENG * EKPO-UNTTO)/100

12 Deviation Quantity $\sum$ EKBE-MENGE- EKPO-MENG

13 Check Quantity commitment GR QTY >= Under Tol. and GR QTY <= Over Tol.

13 %Diff Qty = (Deviation Quantity/EKPO-MENGE) *100

14 Calculate On-time Delivery = Date between EKET-EINDT – EKBE-BUDAT

15 Quality Score = QAVE-QKENNZAHL

// Perform feature Avg.Quantity, Avg.Quality, Avg.Delivery On-time for Supplier performance

16 Avg. GR in Full = If Diff QTY = 0 then 1 else 0

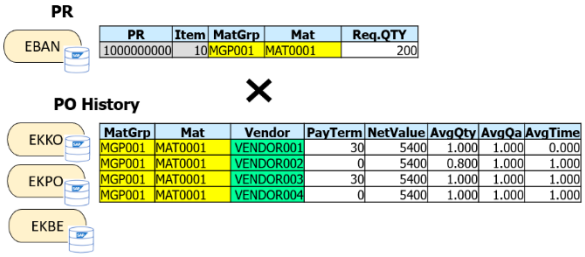
17 Avg.Delivery = If Diff On-time = 0 and <= 5 then 1 else 0

18 Avg.Quality = If Score = 100 then 1 else 0

End

End

การเตรียมข้อมูลสำหรับข้อมูลทดสอบโดยนำข้อมูลใบขอซื้อสินค้า (Purchase Requisition) และข้อมูลประสิทธิภาพของซัพพลายเออร์ในอดีต (Performance history) เพื่อสร้างข้อมูลทำหรับใช้ทดสอบดังรูปที่ 6



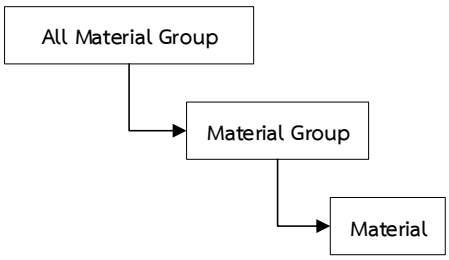
รูปที่ 6 ความสัมพันธ์ของข้อมูลสำหรับทดสอบ

ตารางที่ 6 ข้อมูลสำหรับข้อมูลสำหรับทดสอบ

Feature	Feature Description
Vendor	รหัสผู้ขาย
Material Group	รหัสกลุ่มสินค้า
Material Number	รหัสสินค้า
Payment Term	เงื่อนไขการชำระเงิน
Open Qty	จำนวนสินค้าจากใบสั่งซื้อสินค้า
Net Value	จำนวนเงิน
Avg. Quantity	ค่าเฉลี่ย การส่งสินค้าครบจำนวน
Avg. Delivey	ค่าเฉลี่ย การส่งสินค้าตรงเวลา
Avg. Quality	ค่าเฉลี่ย คุณภาพสินค้า

3.3 สร้างตัวแบบการเรียนรู้ของเครื่องจักรแบบมีผู้สอน

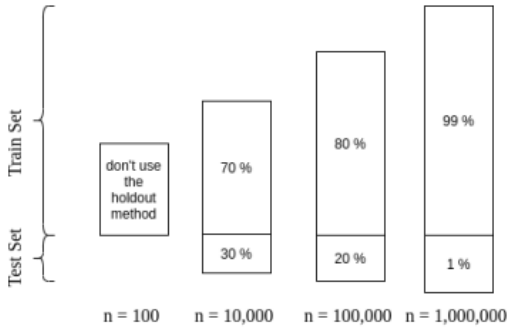
ระดับการวิเคราะห์สามารถแบ่งออกเป็น 3 ระดับ



รูปที่ 7 ระดับการวิเคราะห์ในงานวิจัยครั้งนี้

1. วิเคราะห์ระดับภาพรวมทุกกลุ่มสินค้า (All Material Group) เหมาะสำหรับมีผู้ขายที่มีความสามารถในการขายสินค้าได้หลายกลุ่มสินค้า
2. วิเคราะห์ระดับกลุ่มสินค้า (Material Group) เหมาะสำหรับการประเมินเฉพาะเจาะจงกลุ่มสินค้า
3. วิเคราะห์ระดับสินค้า (Material) วิเคราะห์ระดับสินค้า มีความเฉพาะเจาะจงเป็นรายสินค้า

การสร้างแบบจำลองสำหรับระดับการวิเคราะห์ภาพรวมทุกกลุ่มสินค้า (All Material) แบบจำลองแยกประเภทได้แก่ Naïve bay, k-nearest neighbors(k-NN), Support Vector Machine(SVM), Logistic Regression, Adaboost, Decision tree, Random Forest และ การสร้างแบบจำลองสำหรับระดับการวิเคราะห์กลุ่มสินค้า(Material group) แบบจำลองแยกประเภทได้แก่ ได้แก่ Naïve bay, k-nearest neighbors(k-NN), Support Vector Machine(SVM), Logistic Regression(LR), Adaboost, Decision tree, Random Forest, Stacking model SVM+DT และ Stacking model SVM+LR จำนวนข้อมูล 23,682 แถว โดยแบ่งข้อมูลสำหรับเรียนรู้และข้อมูลสำหรับการทดสอบ



ในอัตราส่วน 70:30 [8-10] และทำการทดสอบแบบ Cross Validation เพื่อแบ่งข้อมูลเป็นชุด ๆ สำหรับการทดสอบแบบจำลองกับทุก ๆ ชุดของข้อมูล

รูปที่ 8 Splitting a Dataset into Train and Test Sets (https://www.baeldung.com/cs/train-test-datasets-ratio)

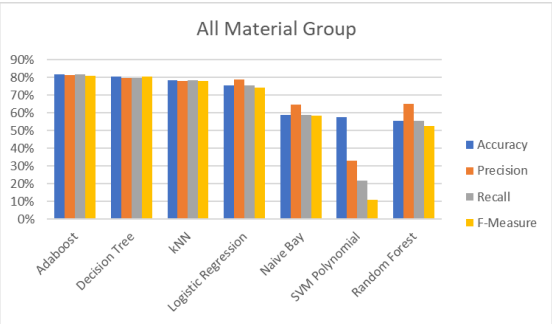
4. ผลการวิเคราะห์และอภิปรายผล

ผลการทดสอบแบบจำลองแยกประเภทโดยการเปรียบเทียบความแม่นยำของแต่ละแบบจำลอง ในระดับภาพรวมทุกกลุ่มสินค้า (All Material Group) และระดับการวิเคราะห์กลุ่มสินค้า (Material group) พบว่าแบบจำลองที่มีความแม่นยำสูงสุดได้แก่ Adaboost Classifier ทั้งสองระดับการวิเคราะห์มีความแม่นยำถึง 81.6% และ 67.6% ตามลำดับ แบบจำลองที่มีความแม่นยำรองลงมา ได้แก่ Decision Tree ที่มีความแม่นยำ 80.3% และ 68.7% โดยกำหนดให้แบบจำลองแยกประเภท Naïve bay เป็นแบบจำลองแยกประเภทพื้นฐาน (baseline model) ในการเปรียบเทียบประสิทธิภาพของ

แบบจำลองที่มีความแม่นยำอยู่ที่ 64.7% และ 55.6% ตามลำดับ ซึ่งแบบจำลองแยกประเภท Adaboost Classifier ทั้งสองแบบจำลองมีความแม่นยำในระดับที่สูงกว่ากว่าแบบจำลองแยกประเภทพื้นฐาน(baseline model) ดังตารางที่ 7

ตารางที่ 7 เปรียบเทียบตัวชี้วัดประสิทธิภาพของแบบจำลองระดับภาพรวมทุกกลุ่มสินค้า (All material group)

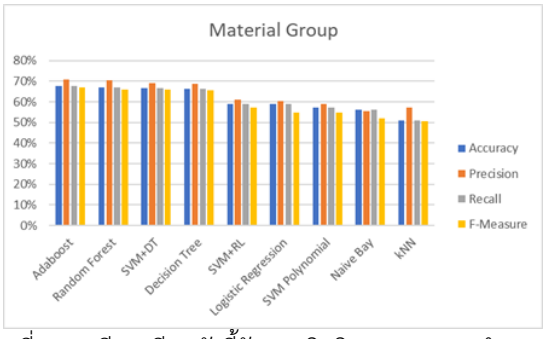
Algorithm (CV=10)	Accuracy	Precision	Recall	F-Measure
Naive (Bay)	0.586	0.647	0.586	0.585
kNN K=8	0.782	0.781	0.782	0.780
SVM Polynomial	0.574	0.329	0.216	0.107
Logistic Regression	0.756	0.786	0.756	0.741
Adaboost	**0.816**	**0.813**	**0.816**	**0.810**
Decision Tree	0.803	0.796	0.797	0.803
Random Forest	0.553	0.652	0.553	0.527



รูปที่ 9 เปรียบเทียบตัวชี้วัดประสิทธิภาพของแบบจำลองระดับภาพรวมทุกกลุ่มสินค้า (All material group)

ตารางที่ 8 เปรียบเทียบตัวชี้วัดประสิทธิภาพของแบบจำลองระดับกลุ่มสินค้า (Material Group)

Algorithm (CV=10)	Accuracy	Precision	Recall	F-Measure
Naive Bay	0.563	0.556	0.563	0.521
kNN K=8	0.509	0.572	0.509	0.506
Decision Tree	0.664	0.687	0.664	0.657
Random Forest	0.669	0.705	0.669	0.660
Adaboost	0.676	0.709	0.676	0.521
SVM Polynomial	0.573	0.591	0.573	0.548
Logistic Regression	0.589	0.605	0.589	0.548
SVM+DT	0.665	0.690	0.665	0.658
SVM+LR	0.590	0.610	0.590	0.571



รูปที่ 10 เปรียบเทียบตัวชี้วัดประสิทธิภาพของแบบจำลองระดับกลุ่มสินค้า (Material group)

5. สรุปผลการการวิจัย อภิปรายผล และข้อเสนอแนะ

ผลการวิจัยในการพัฒนากรอบแนวคิดการจำแนกประเภทซัพพลายเออร์สำหรับปัญหาการประเมินซัพพลายเออร์ในระบบ SAP ERP สามารถเป็นแนวทางการดึงข้อมูล (Extract) และแปลงข้อมูล (Transform) ให้อยู่ในรูปแบบที่มีความเหมาะสมสำหรับนำไปพัฒนาแบบจำลองในการแยกประเภทประสิทธิภาพของซัพพลายเออร์ได้

จากผลการทดลองด้วยการเปรียบเทียบแบบจำลองแยกประเภท พบว่าแบบจำลอง Adaboost Classifier มีความแม่นยำ (Accuracy) สูงสุดเมื่อเทียบกับแบบจำลองพื้นฐาน (baseline model) และแบบจำลองแยกประเภทอื่น ๆ ทั้งนี้ยังมีปัจจัยที่มีผลสำหรับการพัฒนากรอบแนวคิดการจำแนกประเภทซัพพลายเออร์สำหรับปัญหาการประเมินซัพพลายเออร์ในระบบ SAP ERP แบ่งออกเป็น 3 ปัจจัยดังนี้

1. การกำหนดกระบวนการขั้นตอนการทำงาน (Business Process) ที่ชัดเจน เพื่อให้สามารถบันทึกข้อมูลที่เกี่ยวข้องในการประเมินซัพพลายเออร์อย่างมีคุณภาพ

2. การบันทึกข้อมูลหรือทำรายการในระบบ SAP ERP ที่ถูกต้องหรือตามกระบวนการทำงานที่กำหนดไว้

3. บุคลากรที่ทำงานเกี่ยวข้องและสามารถนำข้อมูลไปใช้ในการตัดสินใจ

จากสามปัจจัยดังกล่าวข้างต้น หากได้ทำการเก็บรวบรวมข้อมูลได้จำนวนมากขึ้น และตรวจสอบความถูกต้องของข้อมูลที่ดีมากยิ่งขึ้นในอนาคตแล้ว ตัวแบบการประเมินซัพพลายเออร์ที่พัฒนาขึ้น น่าจะมีโอกาสปรับปรุงความแม่นยำการทำนายได้มากยิ่งขึ้นตามไปด้วย ทั้งนี้ทางผู้วิจัยจะนำไปศึกษาและพัฒนาต่อยอดในลำดับถัดไป

เอกสารอ้างอิง

[1] วิโรจน์ ตันติภัทโร, “การคัดเลือกซัพพลายเออร์ด้วยเทคนิคการวิเคราะห์เส้นกรอบล้อมข้อมูล,” *วารสารวิศวกรรมศาสตร์ มหาวิทยาลัยเอเชียอาคเนย์.*, ปีที่ 2, ฉบับที่ 4, น. 1-16, 2553.

[2] ศุภลักษณ์ ใจสูง, “การคัดเลือกผู้ให้บริการโลจิสติกส์ของบริษัท ฮานา ไมโครอิเล็คโทรนิค จำกัด (มหาชน) โดยใช้กระบวนการตัดสินใจแบบวิเคราะห์ลำดับชั้น (AHP),” *วารสารบริหารธุรกิจ มหาวิทยาลัยธรรมศาสตร์.*, ปีที่ 35, ฉบับที่ 134, น. 65-89, 2555.

[3] สุรพงค์ เอื้อวัฒนามงคล, การทำเหมืองข้อมูล. พิมพ์ครั้งที่ 2., โรงพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย: สำนักพิมพ์สถาบันบัณฑิตพัฒนบริหารศาสตร์. 2561.

[4] อัครนันท์ พงศธรวิวัฒน์, จรุงรัตน์ วรรณสิทธิ์ และชวลิต จีนอนันต์, “ตัวแบบการตัดสินใจเชิงภาษาแบบ 2-Tuple สำหรับปัญหาการตัดสินใจแบบหลายเงื่อนไข กรณีศึกษาการ คัดเลือกผู้ให้บริการโลจิสติกส์สำหรับอุตสาหกรรมเบเกอรี่,” *Journal of Applied Statistics and Information Technology.*, ปีที่ 4, ฉบับที่ 2, น. 31-45, 2563

[5] A. Abdulla, “Weighting the key features affecting supplier selection using machine learning techniques,” Faculty of Engineering, Misurata University, Libya, 2019.

[6] F. Alireza, A. Atefeh, A. Jurgita, Y. Morterza, “Nonlinear genetic-based model for supplier selection: A comparative study, ”*Technological and Economic Development of Economy.*, vol. 23, no. 1, pp. 178-195, 2017.

[7] H. Taherdoost and A. Brard, “Analyzing the Process of Supplier Selection Criteria and Method,” *Procedia Manufacturing.*, vol. 32, pp. 1024 – 1034, 2019.

[8] I. M. Cavalcante, “A supervised machine learning approach to data-driven simulation of resilient supplier selection in digital manufacturing,” *International Journal of Information Management.*, vol. 49, pp. 86–97, 2019.

[9] K. Thakkar, J. Shah, R. Prabhakar, A. Joshi. “AHP and Machine learning techniques for wine recommendation,” *International of computer science and information technologies.*, vol.7, no. 5, pp. 2349-2352, 2016.

[10] M. Kohli. “Supplier Evaluation Model on SAP ERP Application using Machine Learning Algorithms,” *International Journal of Engineering and Technology.*, vol. 7, pp. 306-311, 2017.

[11] Anonymous, “Orange Data Mining,” [Online]. Available: https://orangedatamining.com/. [Accessed 1 December 2020].

[12] N. Pongsathornwiwat, V. N. Huynh, and C. Jeenanunta, “Developing evaluation criteria for partner selection in tourism supply chain networks,” *International Journal of Knowledge and Systems Science.*, vol. 8, no. 1, pp. 39-52, 2017.

[13] N. Pongsathornwiwat, V. N. Huynh, T. Theeramunkong and C. Jeenananta, “A linguistic partner evaluation model in tourism supply chain networks” in *2016 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, 2016.

[14] R. Dingledine and R. Lethin. “Use of Support Vector Machines for Supplier Performance Modeling,” [Online]. Available: svm.pdf (mit.edu). [Accessed 1 December 2020].

[15] “Splitting a Dataset into Train and Test Sets,” [Online]. Available: https://www.baeldung .com/cs/train-test-datasets-ratio. [Accessed 6 February 2021].

[16] V. H. Wilson, A. N. S. Prasad, A. Shankharan, S. Kapoor, J. A. Rajan “Ranking of supplier performance using machine learning algorithm of random forest,” *International Journal of Advanced Research in Engineer and Technology.*, vol. 11, pp. 299-308, 2020.